建议将“Theory of Mind”译为“心理揣测”

赵宇轩[1] 曾毅[1] 秦裕林[2]

（[1]中国科学院自动化研究所，北京 100190）

（[2]上海交通大学凯原法学院，上海 200030）

摘要 “theory of mind”指认识自己和他人的心理状态（目的，意图，注意，信念，知识，欲望，情绪等）的能力。它不仅是心理学和神经科学领域的重要分支，而且在人文社会科学方面有着广泛的应用，同时也是人工智能领域一个极具潜力的发展方向。在中文语境下其含义已经达成共识，但其中文译文尚未达成完全的统一。本研究以“theory of mind”为关键词，在中国知网数据库的学术期刊子库进行检索，经筛选后获得 421 篇符合条件的中文文献，对其分析确定“theory of mind”译文的发展及使用情况。基于“theory of mind”在研究与应用领域实际上指的是心理揣测的状况，以及英语的“theory”翻译为“揣测”可行性及依据，最终建议以“心理揣测”作为“theory of mind”中文译文。

关键词 心理揣测，心理理论，知网数据库

分类号 B84

1 引言

Theory of Mind（ToM）指认识自己和他人的心理状态（目的，意图，注意，信念，知识，欲望，情绪等）的能力。据此，人们可以解释和预测他人的心态和行为，从而能够理解、配合、支持和协助其他社会成员，或者在竞争状态下，干扰竞争对手的行为（Ho et al., 2022）。由于 ToM 涉及人类的基本社会认知功能，不仅哲学上有笛卡尔（1596~1650）所试图解决的“他心问题”（沈学君, 2021），甚至早于他约 250 年的罗贯中（约 1330 年~约 1388 年）就在文学作品中生动地描述了 ToM 思想在军事上的应用：诸葛亮判断在司马懿的心中“诸葛一生唯谨慎”，才敢用“空城计”。事实上这里涉及到二阶 ToM：诸葛亮揣测（司马懿揣测（诸葛亮不会冒险用兵））。但是用实验的方法研究 ToM 只有 40 多年的历史。最早是 Premack 和 Woodruff （1978）以黑猩猩（而不是人类）作被试的研究，Theory of Mind 这个（不那么好的）名字也是他们起的。他们的试验方法也很有趣：让他们实验室 14 岁的黑猩猩 Sarah 看一段 30 秒钟的视频，该视频显示一个人类主角遇到了一个难题（例如举起了手臂，但是够不着天花板上的香蕉）。然后给 Sarah 两张照片，一张照片的画面是问题的解（这个人踩上一个箱子），另一个的画面与问题无关。每个任务重复 4 次。对于三个仅需一步（例如，踩上一个箱子）或两步便能解决的问题，Sarah 能 100%选择问题的解（但是对于一个需要三步解决的问题，而问题解答的照片只显示了第一步动

1中国科学院战略性先导科技专项（编号：XDB32070100）；科技部新一代人工智能重大项目（编号：2020AAA0104305）；北京市类脑计算研究专项（编号：Z181100001518006）资助。
通讯作者：曾毅，yi.zeng@ia.ac.cn

作的任务，Sarah 只是凭机会，50%选择问题的解）。据此，他们认为，黑猩猩至少具有两种 ToM 的能力：推断别人的动机与目的；推断别人的知识与信念。相继的研究除了验证了他们的发现以外，也发现大猩猩的 ToM 不完全，例如他们不能完成“错误信念（false belief）”任务（Call & Tomasello, 2008）。“错误信念”任务最早由 Wimmer 和 Perner（1983）开发用于发展心理学的研究，后来被 Baron-Cohen 等(1985)应用于自闭症研究。“错误信念”任务目前有不同的版本。用得比较普遍的是 Baron-Cohen 等(1985)引进的 Sally 和 Anne 版本：Sally 把一个玻璃球放进自己的篮子后，离开了房间。Anne 在 Sally 离开后，把那颗玻璃球拿出来，放入了 Anne 自己的盒子里，离开了房间。然后提问观察了整个过程的儿童被试：当 Sally 回来的时候，她会在哪里去找玻璃球？正确的答案是去 Sally 自己的篮子里找玻璃球，因为她不知道玻璃球被拿走了。完成这个任务的被试需要知道别人有别人的信念，而且别人的信念可能和自己的不同。一般儿童到 4~5 岁时才开始发展出这种能力（Wimmer and Perner, 1983）。这是一种一阶信念。儿童在 5~10 岁能够发展出二阶信念（例如，“X 认为 Y 感觉......”）（Perner and Wimmer，1985），然后发展出三阶信念（例如，“X 认为 Y 假设 Z 打算......”）等。但是这些只涉及到 ToM 的认知成分（推断他人知识，意图和信念的能力）。ToM 还有情感成分（推断他人情绪状态和感受的能力），他的发展过程与认知成分不同。面部表情识别和欲望在情绪中的作用大约在 2-4 岁左右发展；情绪外部原因的识别以及对信念和记忆在情绪中作用的理解在 5-6 岁左右发展；区分感受和表达情绪的能力在 6-7 岁左右发展；8-9 岁左右就会形成对情绪调节策略的认识（综述参见 Raimo et al., 2022）。认知和情感心理状态的表征涉及的神经回路开始相同，然后走向不同的路径。他们都是在颞顶交界处（TPJ）形成，随后通过颞上沟（STS）或楔前/后扣带复合体（PCun /PCC）到达不同的边缘-旁边缘（limbic–paralimbic）区域，在那里确定是认知或情感。然后，认知心理状态的表征涉及背侧回路（背侧颞极（dTP）、背侧前扣带皮层（dACC）、背内侧前额叶皮层（dmPFC）和背外侧前额叶皮层 （dlPFC）等），而情感心理状态的表征涉及腹侧回路（腹侧纹状体、杏仁核、腹侧颞极（vTP）、腹侧前扣带皮层（vACC）、眶额叶皮层（OFC）、腹内侧前额叶皮层（vmPFC）、和下外侧额叶皮层 （ilFC）等）（Gabriel et al.，2021）。

推断和预测是为了更好地行动。近年来的研究发现 ToM 与执行功能（executive function）密切相关。执行功能有“冷的（cold）”认知与“热的（hot）”情感两个成分。认知成分（cognitive component）包括抑制，工作记忆，规划等认知过程，涉及在情绪中立的情况下做出判断的任务（如威斯康星州卡片分类测试）；情感成分（affective component）包括延迟满足的能力，情感决策等情感过程，涉及充满情感情况下的推理和动机调节任务（例如爱荷华赌博任务）。ToM 认知成分的发展与执行功能的认知成分相关，ToM 情感成分的发展与执行功能的情感成分相关（Longobardi et al., 2022）。从近年来的这

些研究进展来看，在关于 ToM 三种常见机制（模拟论，先天模块论与理论论）的争论中，比较不支持理论论。

由于理解他人是人进行社会行为的基础，这不仅引起认知、社会、发展、比较、临床等心理学与神经科学众多领域的兴趣，而且引起了人文与社会科学众多领域的兴趣。例如，对于 76 项研究（包括 6,432

2 至 12 岁的儿童）的元分析表明，ToM 评分较高的儿童在同时测量的亲社会行为中也获得了更高的分数（r = 0.19）。能够明确考虑他人的想法和感受与儿童亲社会行为的倾向有关。ToM 评分的这种关联对于亲社会行为的不同亚型（帮助，合作，安慰）都很明显。需要识别他人认知的 ToM 和需要识别他人情绪的 ToM 与亲社会行为的关联程度相似（Imuta et al., 2016）。然而另一个对于 81 项研究（包括 7,826

2 至 14 岁的儿童）的元分析也发现，ToM 与儿童说谎行为正相关（r = 0.23）（Lee and Imuta, 2021）。这就突显了儿童教育的重要性。有趣的是，由于理解他人是人类社会行为的基础，ToM 的研究也引起了考古人类学研究者的兴趣。例如 Ando（2016）提供了尼安德特人的工作记忆（执行功能）和 ToM（理解同伴，与同伴协作的能力）比我们智人低的证据。

为了人类的根本利益，我们不仅要警惕和阻止可能有害于人类的人工智能技术，而且要倡导具有人类伦理的人工智能技术的发展。为此需要开发具有理解人类行为的能力，从而能在合适的时侯以合适的方式帮助人类的人工智能技术。例如中佛罗里达大学的 Williams 等人（2022）提出的开发具有人工 ToM 的人工社会智能的设想。斯德哥尔摩经济学院的 Soderlund（2022）考察了具有 ToM 能力的家庭服务机器人可能带来的效应。埃塞克斯大学的 Bianco 等人（2019）提出了适应性心理揣测系统在人机交互中的优势，如信念理解、主动行为、主动感知和学习，增强人与机器人的社会交互。在心理揣测计算建模方面，大致可以分为基于贝叶斯的心理揣测模型、基于深度学习的心理揣测模型、包含连接主义建模、认知架构设计等其他方法的心理揣测模型、以及基于脑启发的心理揣测模型。

基于贝叶斯的心理揣测模型是最具代表性的一类心理揣测模型，以麻省理工学院为代表的科研院校在基于贝叶斯的心理揣测模型方面取得了一系列的代表性成果。麻省理工学院的 Goodman 等人（2006）建立了两个贝叶斯模型，两个模型都支持预测和解释。在简单模型中，Sally 的信念只与玩具的位置相关，而在复杂模型中，Sally 的信念不仅与玩具位置相关还与她对玩具的视觉感知相关，即 Sally 是否能看到玩具移动。这一区别使得简单模型在错误信念任务中失败，而复杂模型会成功。麻省理工学院的 Baker 等人（2017）提出了一个贝叶斯心理揣测模型，该模型可以根据智能体在空间中的移动方式，来推断它的信念、期望和知觉。在两个心理学实验中，该模型获得了和人类被试相似的实验结果。实验结果表明，贝叶斯心理揣测模型可以根据他人的行为揣测他人的信念、期望和知觉，以及用他

人的想法和行为揣测环境的状态。麻省理工学院的 Lee 等人（2019）定义了一个用于人机交互的非语言交流的双重计算框架。他们使用贝叶斯心理揣测方法来模拟讲故事时的交互作用。讲述者利用声音线索来影响和推断听者的注意状态，将其作为一个部分可观测马尔可夫决策规划问题进行计算。听者通过自己的反应传达注意力，将其作为一个动态贝叶斯网络计算。通过人机交互实验证明模型在注意力识别和传达的有效性。爱丁堡大学的 Patacchiola 和 Cangelosi（2020）提出了一种基于信任和心理揣测的发展认知架构，该架构受心理和生物学的启发，由演员-评论家框架和贝叶斯网络组成，这些模块分别对应于大脑中用于心理揣测的脑区。最后，他们用 iCub 仿人机器人进行了两个心理学实验，结果与儿童的实验数据一致，有助于揭示儿童和机器人基于信任的学习机制。

受益于深度学习的飞速发展，基于深度学习的心理揣测模型也取得了很大进展。Google DeepMInd 团队的 Rabinowitz 等人（2018）设计了一个 ToM 神经网络来学习如何通过元学习对其他智能体进行建模。他们构建了一个能够收集智能体行为轨迹的观察者，其目标是预测其他智能体的未来行为。他们将提出的 ToM-net 模型应用于简单的网格环境中，结果表明观察者可以有效地为智能体建模并通过 Sally-Anne 测试。而观察者自身不需要执行任何动作。加利福尼亚大学洛杉矶分校的 Akula 等人（2022）基于心理揣测的思想提出了一个新的可解释 AI 框架 CX-ToM，用于解释深度卷积神经网络做出的决策。该模型可以显式的建模人类用户的意图、人类用户对机器的理解，以及机器对人类用户的理解，通过人类用户和机器之间的多轮交互，提高模型的可解释性，并增加人类对模型的信任。

另外还有一些从连接主义建模、认知架构设计等方面构建心理揣测模型的研究。麦吉尔大学的 Berthiaume 等人（2013）提出了一个连接主义模型来模拟错误信念任务。通过增加隐藏层神经元来提高模型的计算能力，该模型可以成功模拟错误信念任务由失败到成功这一转变，他们认为，这种转变的根源不在于对信念的理解，而是由抑制自身信念处理资源的增加导致的。图卢兹大学的 Milliez 等人（2014）提出了一个时空推理系统 SPARK，借助该系统，机器人可以以更加自然的方式实现有效地沟通和互动。该系统可以使机器人通过 Sally-Anne 测试，并在对话消歧方面表现良好。西英格兰大学的 Winfield（2018）基于内部模拟模型提出了一个心理揣测模型，并部署在 NAO 机器人上，该模型可以在内部模拟机器人下一个可能的动作，从而预测这些动作对自己和其他个体可能产生的后果，对增强机器人的社会交互能力十分重要。西英格兰大学的 Bremner 等人（2019）提出了一个信念-期望-意图模型，其逻辑结构通过记录推理循环和形式化的验证方法促进模型的透明性，通过一系列的实验证明该模型能够做出符合阿西莫夫机器人三定律的正确决策。加州理工学院的 Choudhury 等人（2019）对比了无模型、基于黑箱模型和基于心理揣测的人机交互方法，发现基于心理揣测的人机交互方法是在学习过程中唯一不需要人机交互数据，

并可以根据观察到的人-人交互数据进行训练的方法，相较于另外两个方法，基于心理揣测的方法所需的数据更少，且更加鲁棒。

与认知心理学和脑科学关系更紧密的是脑启发的心理揣测模型。在这个方向上，中科院自动化所曾毅团队取得了一系列研究进展。Zeng 等人（2020）充分借鉴认知心理学、脑科学、神经影像学在心理揣测领域的研究成果，融合心理揣测的多尺度神经基础，即所涉脑区、脑区功能及神经环路，提出并构建脑启发的心理揣测模型。该模型实现了机器人的自我经验学习，并利用自我经验实现对他人信念的揣测及行为预测，使机器人可以通过错误信念任务，获得初步的心理揣测能力。该模型探索了自我经验、相关脑区和脑区间连接的成熟度，特别是抑制控制机制对心理揣测能力的影响，有助于从计算角度揭示心理揣测的神经机制。Zhao 等人（2022a）在此研究的基础上，提出了多脑区协同心理揣测的脉冲神经网络模型。该模型可以区分并对不同类型的智能体进行揣测，并且基于揣测来预判他人未来的安全状态。最后将该模型应用到安全风险任务中，实验证明，具备心理揣测能力的智能体可以帮助他人避免安全风险。心理揣测的心理状态往往是抽象的，不能够表征的。Zhao 等人（2022b）因此不显示构建对他人的心理状态而用网络隐层表征他人的心理状态，进而对他人行为进行预测，帮助提升多智能体合作的性能和效率，并提高智能体在竞争中的竞争力。此外，Zeng 等人（2020）的工作不仅直接贡献于心理揣测计算模型的研究，对心理学和认知科学在心理揣测方面的研究也起到一定的推动作用。格拉斯哥大学的 Roth 等人（2022）认为 Zeng 等人（2020）提出的脑启发的心理揣测模型同认知双重过程方法一致，区分了更自动、快速、更少受控制的过程和更刻意、更缓慢和有意识的过程，并与区分内隐和外显的心理揣测模型一致。 因此，基于脑启发的心理揣测模型（Zeng 等人，2020），特别是该模型提出的四条通路：自我经验学习通路、动机理解通路、自我信念推理通路和他人信念推理通路，以及模型的计算特点，Roth 等人提出了一项新的心理学实验范式，并在 60 名人类被试上进行了该实验，实验结果进一步证明了脑启发心理揣测模型的有效性和合理性，有助于进一步揭示心理揣测的神经机制。两项工作将人工智能和心理学在心理揣测方面的研究紧密结合在了一起，相互促进，相互补足，形成良好循环。

由于以上的原因，ToM 的研究发展得非常迅猛。我们用“Theory of Mind”检索 Google Scholar 得到 651,000 篇文献【注 1】，并且发现，从 1990 年代开始 ToM 方向的论文发表就呈现了加速发展的状态。进入本世纪以后，加速度更是迅速增大。2012 年起每年新发表的关于 ToM 的论文就超过 10,000 篇（图 1）。

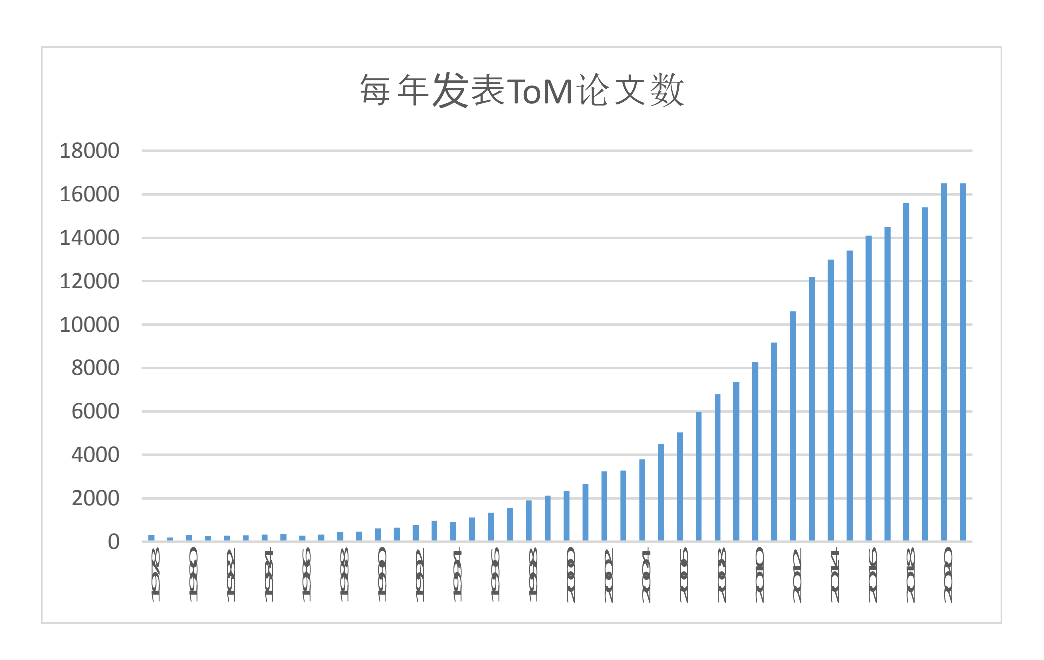


图 1. 1978 年~2021 年，每年发表 ToM 论文数

与这些轰轰烈烈发展的研究所相悖的是其词不达意的命名--“Theory of Mind”。Premack 和 Woodruff（1978）说把这种推断心理状态的系统称为“theory”是因为心理状态是不能直接观察到的，并且这个系统可以用来预测他人的行为（A system of inferences of this kind is properly viewed as a theory because such states are not directly observable, and the system can be used to make predictions about the behavior of others.）。讲英语的人中就有人对这种命名不满（关于 ToM 机制的理论论的存在，正说明了 ToM 并不就等于就是理论）。例如 Apperly（2011）就认为把这种推测别人心理状态的能力直接叫做“mindreading”比较好，并且出版了以“mindreaders”为主标题的关于成年人 ToM 的书（书中的第一句话是：Theory of mind, or “mindreading” as it is termed in this book, is the ability to think about beliefs, desires, knowledge and intentions.）。到了中文，问题就更糟了。或许是以为“theory”只能翻译成“理论”，“Theory of Mind”在很多时候被翻译成了“心理理论”。结果明明是“揣测”，却又说成是“理论”，包括教科书在内，例如：“Theory of mind (mentalizing)/心理理论：反观自我和推测他人想法的能力，推测他人能够理解的内容和在特定环境中的反应与彼此作用。这种能力是人类特有的。”（Gazzaniga 等著，2011 中文版，术语表）。结果给应用与教学带来许多人为造成的不便。事实上，在英语里“theory”有“揣测”的义项。把“Theory of Mind”直接翻译成“心理揣测”就可以避免这种“名不符实”的问题。

本文下面的几节将按照如下顺序讨论这个问题：1. 中文文献中“Theory of Mind”的不同译名及其定义。可以看到，尽管译名有所不同，定义基本上都是“推断”；2. 中文文献检索的方法；3. 中文文献检索的结果；4. “揣测”对应“theory”的讨论，包括英语词典和英-汉词典中关于“theory”的“揣测”的义项的列举和 John Anderson 关于“theory”有两种含义的个人通讯。

最后，综合 Premack 和 Woodruff（1978）意义下使用的“Theory of Mind”的“Theory”在研究与应用中实际是“揣测”的含义和英语中“theory”具有“揣测”的义项，我们建议将 Premack 和 Woodruff（1978）意义下使用的“Theory of Mind”翻译为“心理揣测”,以改变目前的“词不达意”的现象。

2 中文文献中“Theory of Mind”的不同译名及其定义

“Theory of mind” 不仅是心理学和神经科学领域的重要分支，而且在人文社会科学方面有着广泛的应用。在中文论文中，“theory of mind”大多被翻译为“心理理论”，其他还存在如“心理推理能力”、“心灵理论”、“心智理论”、“心理推测能力”、“心理推测”、“心的理论”、“心理推断”、“思维理论”、“想法解读理论”等译文，定义如下所示：

“心理理论”指个体具有将自身及其他个体的行为归因为心理状态的能力, 由此产生的对行为原因的推论组成一个理论系统。将其视为理论有两点理由: 首先, 心理状态不能被直接观察; 其次, 这一推论系统能被用来预测个体行为（王茜等，2000）。

“心理推理能力”指的是个体理解自我和他人的愿望、意图、信念等心理状态, 并依此对行为做出解释和预测的能力（金宇等，2002）。

“心灵理论”指个体所具有的对他人意图、需要、动机、信念、情感、愿望等心理状态进行推测的能力结构系统（张登科等，2011）。

“心智理论”指的是一种与心智有关的能力，包括认为自己和他人具有心理状态的能力和应用心理知识预期自己和他人行为的能力（周统权，徐晶晶，2012）。

“心理推测能力”是个体对自己和他人心理状态 (如需要、信念、意图、感觉等) 的认识, 并由此对相应行为做出因果性解释和预测的能力（王立新等，2003）。

“心理推测”是指个体将独立的心理状态归因到自己或他人，并依据这些心理状态去预测和解释他人行为的能力。心理状态既指个体的目的，意图，知识，信念，思考，怀疑猜测，假装，喜好，也包括知觉和情绪等（焦青，2000）。

“心的理论”是指对他人的心理状态推测判断的能力（徐光兴，2000）。

“心理推断”能力是指儿童是否能认知诸如信念、需要、意愿等心理状态以及这些心理状态与外部世界或外在行为之间的关系（桑标等，1994）。

“思维理论”是把信念、情感和意图归因于人的其他能力（陈鹤三，2011）。

“想法解读理论”是指推测别人想法的能力，并同时运用这种能力来理解别人的说话或行为背后的用意，和预计他们下一步的行动（孙广宇，2006）。

由此可见，在中文语境下，“theory of mind”中“theory”的含义已经达成共识，不是“理论”，而是“揣测”。但其中文译文尚未达成完全的统一。我们通过检索已经公开发表的中文学术期刊论文，确定“theory of mind”中文译文的发展及使用情况，并提出以“心

理揣测”作为其中文译文的可行性及依据。

3 研究方法

3.1 文献来源

文献检索涵盖了中文期刊文献，检索出版日期截止至 2021 年。在中国知网数据库的学术期刊子库进行检索，检索关键词为“theory of mind”，来源类别包括北大核心（北京大学图书馆“中文核心期刊”）、CSSCI（南京大学“中文社会科学引文索引来源期刊”）、CSCD（中国科学院文献情报中心“中国科学引文数据库来源期刊”），共检索获得 544 篇文献。

3.2 文献纳入标准

将检索得到的文献按照以下标准进行筛选：（1）排除哲学、中国文学、宗教、中等教育、中国语言文字、外国语言文字、世界文学、文艺理论学科下的不相关文献；（2）文献的英文题目、英文摘要、英文关键词、正文中至少有一处包含“theory of mind”。

共排除文献 123 篇，其中哲学、文学等学科下不相关文献 83 篇，不含“theory of mind”文献 40 篇。最终保留文献 421 篇，出版时间为 1994 年至 2021 年，其中 1995、1996、1997、1998 年无符合条件的出版文献。

4 研究结果

我们统计了每年符合条件的文献篇数，如图 2 所示。在 2004 至 2014 年间，符合条件的文献篇数保持较高数量，每年符合条件的文献篇数基本全在 20 篇以上。

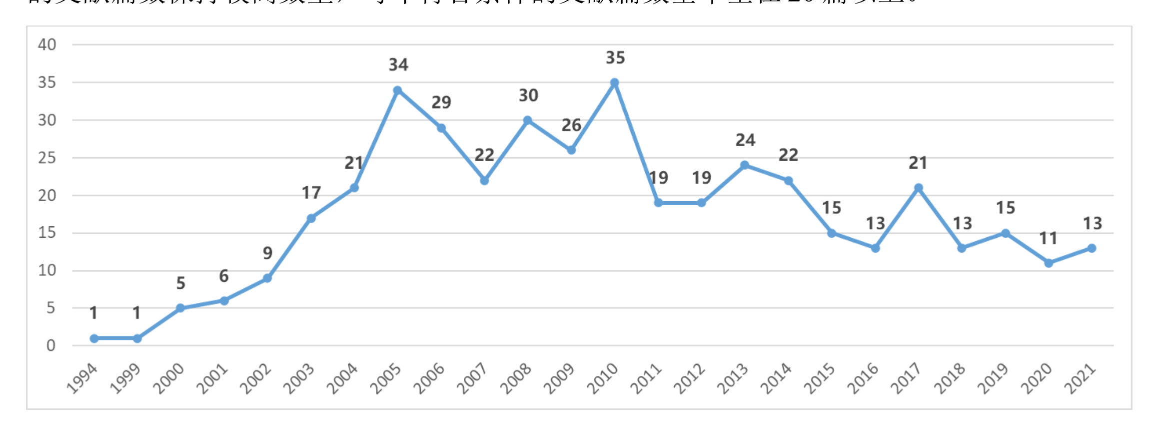


图 2 每年符合条件的文献篇数

421 篇文献中将“theory of mind”翻译为“心理理论”、“心理推理能力”、“心灵理论”、“心智理论”、“心理推测能力”、“心理推测”、“心的理论”、“心理推断”“思维理论”、“想法解读理论”十种。

我们统计了每年每种译文情况的占比，结果如图 所示。（1）在 2003 年之前，符合条件的相关文献较少，“theory of mind”一词的翻译尚未统一，译为“心理理论”、“心灵理论”、“心理推测能力”、“心理推测”、“心的理论”、“心理推断”，共六种，其中译为“心理理论”占比较大。（2）在 2004 至 2014 年间，“theory of mind”主要被译为

“心理理论”，占比大多数在 95%以上。（3）在 2015 年后，“theory of mind”被译为“心理理论”、“心理推理能力”、“心灵理论”、“心智理论”、“心理推测能力”，共五种，译为“心理理论”的情况占比略有降低。

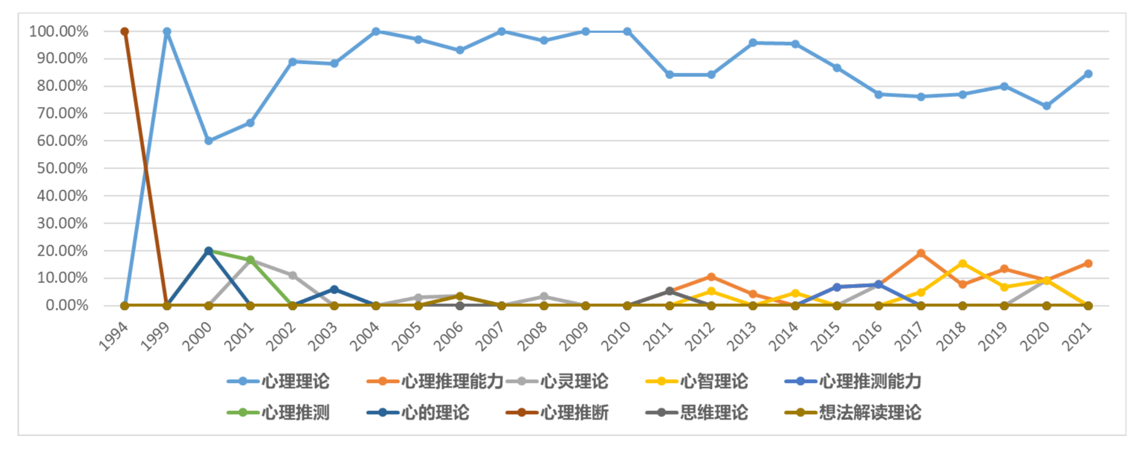


图 3 每年每种译文情况占比

421 篇文献中，按译文总数从高到低排序，结果如表格 1 所示。从总数上来看，“theory of mind”大多被译为“心理理论”，占比高达 90%。

表格 1 “theory of mind”译文一览表

译文	总数	占比	最早时间	最晚时间
心理理论	380	90. 26%	1994	2021
心理推理能力	16	3. 80%	2011	2021
心灵理论	8	1. 90%	2001	2020
心智理论	7	1. 66%	2012	2020
心理推测能力	3	0. 71%	2003	2006
心理推测	2	0. 48%	2000	2001
心的理论	2	0. 48%	2000	2003
心理推断	1	0. 24%	1994	1994
思维理论	1	0. 24%	2011	2011
想法解读理论	1	0. 24%	2006	2006

此外，我们统计了文献篇数大于十篇的来源期刊对“theory of mind”一词翻译的分布情况，如表格 2 所示。在这些期刊中，除《中国心理卫生杂志》将“theory of mind”译为“心理推理能力外”大多数期刊都将其翻译为“心理理论”。

表格 2 文献篇数大于十篇的来源期刊译文分布情况

译文	总数	心理理论	心理推理能力	心灵理论	心智理论	心理推测能力	心理推测	心的理论	心理推断	思维理论	想法解读理论
心理科学	62	60	–	–	–	–	–	1	1	–	–
中国特殊教育	42	38	–	–	–	1	1	1	–	–	1
心理科学进展	38	38	–	–	–	–	–	–	–	–	–
心理发展与教育	37	37	–	–	–	–	–	–	–	–	–
心理学报	27	27	–	–	–	–	–	–	–	–	–
中国临床心理学杂志	28	28	–	–	–	–	–	–	–	–	–
中国心理卫生杂志	26	10	14	1	–	–	1	–	–	–	–
心理学探新	18	18	–	–	–	–	–	–	–	–	–
心理与行为研究	15	15	–	–	–	–	–	–	–	–	–
中华行为医学与脑科学杂志	14	13	1	–	–	–	–	–	–	–	–

学前教育研究	12	12	–	–	–	–	–	–	–	–	–
其他	102	84	1	7	7	2	–	–	–	1	–

最后，我们统计了除“心理理论”外的译文在来源期刊的分布情况，如表格 3 所示。除“心理理论”外的译文在期刊中呈现一定的聚集现象，如“心理推理能力”的译文主要分布在《中国心理卫生杂志》中，约占 92.86%，其他情况的译文约在 50%左右。此外，可见其中大多数期刊为基础医学、文史哲综合、教育等学科专题期刊。

表格 3“心理理论”外的译文在来源期刊的分布情况

译文	**总数**	**来源期刊**
心理推理能力	16	中国心理卫生杂志，14 中华行为医学与脑科学杂志，1 中国老年学杂志，1
心灵理论	8	中国神经精神疾病杂志，4 自然辩证法通讯，1 陕西师范大学学报(哲学社会科学版)，1 中国心理卫生杂志，1 华中师范大学学报(人文社会科学版)，1
心智理论	7	自然辩证法通讯，3 西南大学学报(社会科学版)，1 西安交通大学学报(社会科学版)，1 中国现代医学杂志，1 外语教学，1
心理推测能力	3	中华精神科杂志，2 中国特殊教育，1
心理推测	2	中国心理卫生杂志，1 中国特殊教育，1
心的理论	2	中国特殊教育，1 心理科学，1
心理推断	1	心理科学，1
思维理论	1	外语研究，1
想法解读理论	1	中国特殊教育，1

5 “揣测”对应“theory”的讨论

我们认为“theory of mind”的译文大致经历了三个阶段：2003 年之前的各抒己见、2004 至 2014 年的达成共识，再到 2015 年之后的同中求异，如图 所示，括号内为对应译文的出现次数及所占比例。可见，在各个阶段，研究人员在翻译“theory of mind”时都试图表现出该词推理、推测的含义，这也同样是“theory of mind”重要特点之一。但“theory of mind”的主流译文“心理理论”并不能直接简单的表达出这类含义。在《辞海（第六版彩图本）》（辞海（第六版彩图本），2009）中，“理论”的定义包括“概念、原理的体系”。

在《新华字典（第 10 版）》（新华字典（第 10 版，2004）中，“理论”指的是“人们从实践中概括出来又在实践中证明了的关于自然界和人类社会的规律性的系统的认识。”

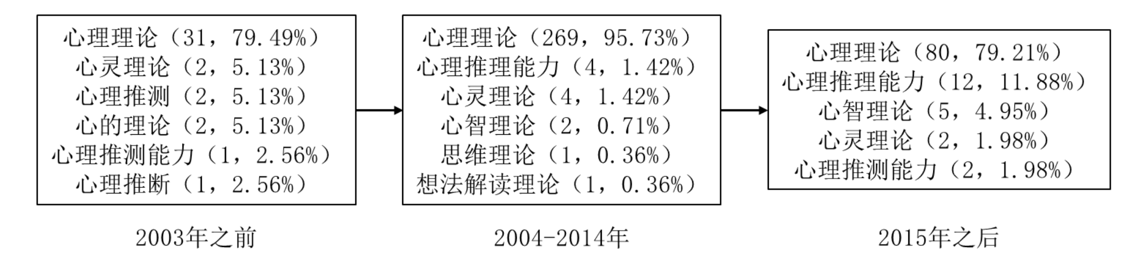


图 4 “theory of mind”译文发展经历

为了促进多科学的交叉融合，使译文更贴切英文含义，并保证译文的直观、简练、易读，我们建议可以将“theory of mind”译为“心理揣测”一词，“揣测”对应“theory”的依据如下：

(1) 在《Oxford Advanced Learner’s Dictionary 8th edition》（Oxford Advanced Learner's Dictionary 8th edition，2010）中，

- “theory”存在如下定义：① a formal set of ideas that is intended to explain why sth happens or exists、② an opinion or idea that somebody believes is true but that is not proved；

(2) 在《Oxford English Dictionary》（Oxford English Dictionary，1989）中，

- “theory”存在如下定义：① Mental view、② In loose or general sense: A hypothesis proposed as an explanation; hence, a mere hypothesis, speculation, conjecture; an idea or set of ideas about something; an individual view or notion、③ A scheme or system of ideas or statements held as an explanation or account of a group of facts or phenomena; a hypothesis that has been confirmed or established by observation or experiment, and is propounded or accepted as accounting for the known facts; a statement of what are held to be the general laws, principles, or causes of something known or observed；

(3) 在《牛津高阶英汉双解词典（第 7 版）》（牛津高阶英汉双解词典：第 7 版，2009）中，

- “theory”存在如下定义：an opinion or idea that somebody believes is true but that is not proved（未证明的）意见；看法；推测”；

(4) 在《英汉大词典（第二版）》（英汉大词典（第二版，2007）中，

- “theory”可被翻译为“意见，看法；推测，揣度”。

(5) 在《牛津当代百科大辞典》（牛津当代百科大辞典（英汉·英英·彩色·图解），2004）中，

- “theory”可被翻译为：①（解说一连串事实或现象的）理论，学说、②个人的见解，主张；推测，推论、③“语源，希腊语，观察、推测之意”。

(6) 在《新华字典（第 10 版）》（新华字典（第 10 版，2004）中，

- “揣”指的是：估量，衬度。

(7) 在《辞海（第六版彩图本）》（辞海（第六版彩图本），2009）中，

- “揣测”指的是：推测。

(8) 在《现代汉语词典（第 6 版）》（现代汉语词典（第 6 版，2012）中，

- “揣测”指的是：推测；猜测。

(9) 在《现代汉语大词典》（现代汉语大词典，2010）中，

- “揣测”指的是：揣度；推测。

因此，我们认为，从英文单词及中文词语的含义上考虑，“theory of mind”是可以被译为“心理揣测”的，并可以更加直接的表达出“theory of mind”的含义。

本文作者之一曾经用电子邮件向美国科学院院士，卡内基·梅隆大学心理学与计算机科学讲席教授约翰·安德森（John Anderson）求证过英语的 Theory 有两个含义，对应于中文的“理论”（Systematically organized knowledge applicable in a relatively wide variety of circumstances）和“揣测”（An assumption based on limited information or knowledge; a conjecture）”。Theory of Mind 中的 theory 用的是第二含义（揣测）。安德森对此表示同意，并且对中文有两种不同的单词对应于 theory 两种不同的含义表示赞赏。对于第一含义（理论）给出了在《心理学评论》中发表的《An integrated theory of the mind》作为例子。【原文见附录】

6 结论

在本文中，我们首先总结分析了“theory of mind”一词在中文期刊中译文的发展及使用情况，然后提出将其翻译为“心理揣测”一词的可行性及依据，希望借此可以进一步促进多学科的交叉融合，使译文更贴切英文含义，并保证译文的直观、简练、易读。

【附录】

本文作者之一于 2013 年 10 月 11 日 给安德森的电子邮件：

Dear Professor,

The word of ‘Theory’ is usually understood as ‘Systematically organized knowledge applicable in a relatively wide variety of circumstances.’

However, I found from the dictionary, it also has the meaning of ’An assumption based on limited information or knowledge; a conjecture’

These two meanings are corresponded to totally different words in Chinese.

Currently in China, the ‘theory’ in the term of ‘Theory of mind’ is translated as the first meaning (Systematically organized knowledge), but I think it should be with the second meaning (a conjecture). Is my guess correct? How do you think about it?

安德森于 2013 年 10 月 12 日 11：20 回复的电子邮件：

It is interesting that Chinese has difference words for the two senses. I assume you are referring to "theory of mind" as in the developmental and neuroimaging and other recent cognitive science research. I think the term also has an earlier history in philosophy. In any case, the second definition does seem closer to cognitive science research as you suggest. My own sense of the English word "theory", outside of its attachment to "mind", is that it is used with a range of shades of meaning spanning the two you quote.

John R. Anderson
Richard King Mellon Professor of Psychology and Computer Science
Carnegie Mellon University

安德森于 2013 年 10 月 12 日 20：49 的电子邮件：

I just realized we called our Psych Review paper "An integrated theory of the mind". I came up with that title from the earlier philosophical tradition and meant it in the first definition. It would be nice if English had two words -- one I could have used in the title of that paper to keep it separate from this other cognitive science research.

【注 1：2022 年 11 月 7 日检索】

参考文献

陈鹤三. 再论批评话语分析的认知层面——进化心理学对批评话语分析的启示[J]. 外语研究, 2011(4): 23-29.

辞海（第六版彩图本）[M]. 上海: 上海辞书出版社, 2009.

焦青. 孤独症儿童心理推测能力的影响因素的研究[J]. 中国特殊教育, 2000(3): 38-40.

金宇,静进,森永良子,等. 中国日本儿童心理推测能力比较研究[J]. 中国心理卫生杂志, 2002(7): 446-448.

陆谷孙. 英汉大词典（第二版）[M]. 上海: 上海译文出版社, 2007.

牛津当代百科大辞典（英汉·英英·彩色·图解）[M]. 北京: 中国人民大学出版社, 2004: 709.

牛津高阶英汉双解词典：第 7 版[M]. 北京: 商务印书馆, 2009.

桑标,缪小春,陈美珍. 幼儿对心理状态的认识[J]. 心理科学, 1994(6): 328-333, 362, 384.

沈学君.(2021). 笛卡尔他心理论浅析. 上海交通大学学报（哲学社会科学版）,(29)141, 112-119.

孙广宇. 运用“想法解读”理论提高孤独症学生对他人情绪辨别能力的个案研究[J]. 中国特殊教育, 2006(10): 74-78.

王立新,彭聃龄,王培梅. 自闭症认知缺陷的神经机制研究进展[J]. 中国特殊教育, 2003(3): 78-82.

王茜,苏彦捷,刘立惠. 心理理论—— 一个广阔而充满挑战的研究领域[J]. 北京大学学报(自然科学版), 2000(5): 732-738.

现代汉语词典（第6版）[M]. 北京: 商务印书馆, 2012.

现代汉语大词典 [M]. 上海: 汉语大词典出版社, 2010.

新华字典（第10版）[M]. 北京: 商务印书馆, 2004: 363.

徐光兴. 关于自闭症的临床、实验心理学的研究[J]. 心理科学, 2000(1): 38-41, 67-125.

张登科,苏巧荣,张宏卫,等. 局限性脑外伤患者的心理推理能力和执行功能缺陷[J]. 中国心理卫生杂志, 2011, 25(v.25): 549-555.

周统权,徐晶晶. 心智哲学的神经、心理学基础:以心智理论研究为例[J]. 外语教学, 2012, 33(1): 8-15.

Akula, A. R., Wang, K., Liu, C., Saba-Sadiya, S., Lu, H., Todorovic, S., ... & Zhu, S. C. (2022). CX-ToM: Counterfactual explanations with theory-of-mind for enhancing human trust in image recognition models. Iscience, 25(1), 103581.

Ando, J. (2016). Evolutionary locus of the neanderthal between chimpanzees and modern humans: A working memory, theory of mind, and brain developmental, Piagetian perspective. In Terashima, H. & Hewlett, B. S. (ed.) Social Learning and Innovation in Contemporary Hunter-Gatherers, Evolutionary and Ethnographic Perspectives. Springer Japan KK. 293-309.

Apperly, L. (2011). Mindreaders: the cognitive basis of “theory of mind”, Psychology Press， New York, NY.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nature Human Behaviour, 1(4), 1-10.

Baron-Cohen, S., Leslie, A. H., & Frith, U. (1985). Does the autistic child have a “theory of mind”? Cognition, 21, 37-46.

Berthiaume, V. G., Shultz, T. R., & Onishi, K. H. (2013). A constructivist connectionist model of transitions on false-belief tasks. Cognition, 126(3), 441-458.

Bianco, F., & Ognibene, D. (2019). Transferring Adaptive Theory of Mind to social robots: insights from developmental psychology to robotics. Paper presented at the International Conference on Social Robotics.

Bremner, P., Dennis, L. A., Fisher, M., & Winfield, A. F. (2019). On Proactive, Transparent, and Verifiable Ethical Reasoning for Robots. Proceedings of the IEEE, 107(3), 541-561.

Call, J & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. Trends in Cognitive Science. 12(5), 187-192

Choudhury, R., Swamy, G., Hadfield-Menell, D., & Dragan, A. D. (2019). On the utility of model learning in HRI. Paper presented at the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI).

Gabriel, E. T., Oberger, R., Schmoeger, M., Deckert, M., Vockh, S., Auff, E., Willinger, U. (2021). Cognitive and affective Theory of Mind in adolescence: developmental aspects and associated neuropsychological variables. Psychological Research, 85, 533-553.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. 著，周晓林，高定国等译. (2011). 认知神经科学：关于心智的生物学, 中国轻工业出版社. 第 595 页.

Goodman, N. D., Baker, C. L., Bonawitz, E. B., Mansinghka, V. K., Gopnik, A., Wellman, H., ... & Tenenbaum, J. B. (2006, July). Intuitive theories of mind: A rational approach to false belief. In Proceedings of the twenty-eighth annual conference of the cognitive science society (Vol. 6). Vancouver: Cognitive Science Society.

Ho, M. K., Saxe, R. & Cushman F. (2022). Planning with Theory of Mind. Trends in Cognitive Science. 26(11), 959-971.

Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review. Developmental Psychology, 52(8), 1192-1205.

Lee, J. J., Sha, F., & Breazeal, C. (2019). A Bayesian theory of mind approach to nonverbal communication. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 487-496). IEEE.

Lee, J.Y. S., & Imuta, K. (2021). Lying and theory of mind: A meta-analysis. Child Development, 92(2), 536-553.

Longobardi, E., Spataro, P., Morelli, M., & Laghi, F. (2022). Executive function ratings in educational settings: concurrent relations with cognitive and affective theory of mind. Early Child Development and Care, 192(13), 2046-2058.

Milliez, G., Warnier, M., Clodic, A., & Alami, R. (2014). A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management. Paper presented at the Ro-Man: the IEEE International Symposium on Robot and Human Interactive Communication.

Oxford Advanced Learner's Dictionary 8th edition[M]. Oxford: Oxford University Press, 2010.
Oxford English Dictionary[M]. Oxford: Oxford University Press, 1989.

Patacchiola, M., & Cangelosi, A. (2020). A developmental cognitive architecture for trust and theory of mind in humanoid robots. IEEE Transactions on Cybernetics.

Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that. . .” attribution of second-order beliefs by 5- to 1O-year-Old children. Journal of Experimental Child Psychology, 39, 437-471.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? The Behavior and Brain Science.4, 515-526.

Rabinowitz, N. C., Perbet, F., Song, H. F., Zhang. C., Eslami, S. M. A., & Botvinick, M. (2018). Machine Theory of Mind. arXiv:1802.07740 [cs.AI]

Raimo, S., Cropano, M., Roldán-Tapia, M. D., Ammendola, L., Malangone, D., & Santangelo, G. (2022). Cognitive and affective theory of mind across adulthood. Brain Science, 12(899), 1-14.

Roth Marion, Marsella Stacy, & Barsalou Lawrence. (2002, November). Cutting Corners in Theory of Mind. In AAAI Fall Symposium 2022 on Thinking Fast and Slow and Other Cognitive Theories in AI (pp. 1-16).

Soderlund，M. (2022) Service robots with (perceived) theory of mind: An examination of humans’ reactions. Journal of Retailing and Consumer Services, 67(102999), 1-11.

Williams, J., Fiore. S. M., & Jentsch, F. (2022). Supporting artificial social intelligence with theory of mind. Frontiers in Artificial Intelligence, 5, 750763-750763.

Wimmer, H., & Perner, J.（1983）Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception. Cognition, 13, 103-128

Winfield, A. F. (2018). Experiments in artificial theory of mind: From safety to story-telling. Frontiers in Robotics and AI, 5, 75.

Zeng, Y., Zhao, Y., Zhang, T., Zhao, D., Zhao, F., & Lu, E. (2020). A brain-inspired model of theory of mind. Frontiers in Neurorobotics, 14, 60.

Zhao, Z., Lu, E., Zhao, F., Zeng, Y., & Zhao, Y. (2022a). A Brain-Inspired Theory of Mind Spiking Neural Network for Reducing Safety Risks of Other Agents. Frontiers in neuroscience, 446.

Zhao, Z., Zhao, F., Zhao, Y., Zeng, Y., & Sun, Y. (2022b). Brain-Inspired Theory of Mind Spiking Neural Network Elevates Multi-Agent Cooperation and Competition. Available at SSRN 4271099.

Suggested translation of "Theory of Mind" as "心理揣测"

Abstract "Theory of Mind" refers to the ability to understand mental states (purposes, intentions, attention, beliefs, knowledge, desires, emotions, etc.) in oneself and others. It is not only an important branch in the field of psychology and neuroscience, but also has a wide range of applications in humanities and social sciences, as well as a possible development direction for artificial intelligence. In the Chinese context, its meaning has been agreed upon, but its Chinese translation has not yet been fully unified. In this paper, we searched the sub-collection of academic journals in the China National Knowledge Infrastructure database with "theory of mind" as the keyword, and obtained 421 eligible Chinese documents after screening, and analyzed them to determine the development and usage of "theory of mind". The development and usage of the translation of "theory of mind" were analyzed. Based on the fact that "theory of mind" actually refers to the state of mental speculation in the field of research and application, and that the translation of "theory" into "揣测" in English. We propose to use "theory of mind" as the Chinese translation of "心理揣测" based on its feasibility and basis.

Key words: 心理揣测，Theory of Mind，China National Knowledge Infrastructure